The Use of Proteomics Techniques to Identify Potential Markers of Early Stage Colorectal Cancer

By

Donna May O'Neil

A thesis submitted to the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Cancer Sciences

College of Medical and Dental Sciences

The University of Birmingham

Febuary 2011

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Colorectal cancer (CRC) is the third most common cancer and the second most common cause of cancer related deaths in the UK. Research has shown that the five year survival rate for patients if diagnosed at an early stage is 83% however, only 11% of cases are diagnosed at this stage. The aim of this study was to use proteomic approaches to investigate secreted proteins from colorectal cancer cell lines to identify candidate biomarkers for early stage diagnosis. Microvesicles (MVs) are a mixed population of vesicles that are released by a wide range of cells and are thought to play a role in tumour development and progression. Stable Isotope Labeling of Amino Acids in Culture (SILAC) was used to investigate the relative abundance of proteins secreted in MVs released by two cell lines that are used as a model of early tumour progression. This study identified 86 potential candidates that demonstrated increased release and six of these proteins (AGR2, OLFM4, SBP1, HSP90α, HSP90β and CEACAM5) were selected for further investigation by Western blot analysis. These proteins show potential as markers of early stage CRC and would be suitable for further validation in patient serum samples.

Contents

Abstract....2

Acknowledgements....4

Contents....5

List of Figures....7

List of Tables....11

List of Supplementary Material (see enclosed CD)....13

Abbreviations....16

Chapter 1 : Introduction....18

Section 1.1 : Colorectal Cancer and Currently used Biomarkers....18

Section 1.2 : Mass Spectrometry-based Proteomics....30

Section 1.3 : Data analysis for Shotgun Proteomic Approaches for Biomarker Discovery....55

Section 1.4 : Cancer and Biomarker Discovery....67

Section 1.5 : Microvesicles....89

Section 1.6 : Aims and Objectives....99

Chapter 2 : Materials and Methods....101

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify potential biomarkers for the early detection of colorectal cancer using proteomic techniques. 130

Chapter 4 : Optimisation of Conditions Required for Quantitative Proteomic Profiling of MVs Released by RG/C2 and PRG/C2 Cell Lines. 156

Chapter 5 : Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labeling of Amino acids in Culture (SILAC) 187

Chapter 6 : Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression 220

Chapter 7 : Discussion 244

References 254

List of Figures

Chapter 1 : Introduction

Figure 1. 1 : Cancer Research UK Colorectal Cancer Statistics. 20

Figure 1. 2 : Dukes' Staging of CRC. 22

Figure 1. 3 : Tandem MS/MS based shotgun proteomics. 31

Figure 1. 4: Ion Fragmentation of Peptides. 36

Figure 1. 5 : Stable Isotope Labeling by Amino acids in Culture (SILAC) 38

Figure 1. 6 : Examples of Chemical Tags used for Quantitative Proteomics 41

Figure 1. 7 : Ionistion Methods used in Proteomics 46

Figure 1. 8 : Mass Analysers 50

Chapter 2 : Materials and Methods

Figure 2. 1 : LC MS/MS set up used in this study. 116

Figure 2. 2 : Calibration and Tuning Spectra. 118

Figure 2. 3 : Acetonitrile gradients used in this Study. 121

Figure 2. 4 : The TransProteomic Pipeline (TPP). 127

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify possible biomarkers for the early detection of colorectal cancer using proteomic techniques

Figure 3. 1 : Initial 2D-PAGE Analysis of CM collected from the RG/C2 and PRG/C2 cell lines. 132

Figure 3. 2 : 2D-PAGE Analysis of CM collected from the RG/C2 and PRG/C2 cell lines. 133

Figure 3. 3 : 2D-PAGE Analysis of CM collected from the RG/C2 and PRG/C2 cell lines following sample cleanup using the 2D Cleanup Kit. 136

Figure 3. 4 : Investigating the presence of EMMPRIN in RG/C2 and PRG/C2 cells and its release into CM. 140

Figure 3. 5 : Enrichment of MVs prepared from CM derived from RG/C2 cells using Sucrose Density Gradient Ultracentrifugation. 142

Figure 3. 6 : Preparation of an MV fraction from CM derived from RG/C2 cells ($6x10^6$ cells) to attempt to identify EMMPRIN using proteomic techniques. 144

Figure 3. 7 : Investigating the presence of EMMPRIN in MVs from RG/C2 cells ($1.2x10^7$ cells) using proteomic techniques 146

Figure 3. 8 : Investigating the presence of EMMPRIN in MVs from RG/C2 cells ($2.4x10^7$ cells) using proteomic techniques 148

Chapter 4 : Optimisation of Conditions Required for Quantitative Proteomic Profiling of MVs Released by RG/C2 and PRG/C2 Cell Line

Figure 4. 1 : CBB G250 versus Silver staining? 160

Figure 4. 2 : Examples of MS/MS spectra that match the theoretical spectra in the database and are considered to be ‘good hits’ and ‘poor hits’. 166

Figure 4. 3 : Manual Inspection of the peptides from the proteins gi|48425720|pdb|1SUV|B Chain B, Structure Of Human Transferrin Receptor-Transferrin Complex and gi|136378|sp|P02786|TFR1_HUMAN Transferrin receptor protein 1 (TfR1) identified in the NCBI database search results. 177

Chapter 5 : Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labelling of Amino acids in Culture (SILAC)

Figure 5. 1 : SILAC Labeling Efficiency. 188

Figure 5. 2 : First collection of MV's from SILAC labeled RG/C2 and PRG/C2 cells. 190

Figure 5. 3 : Example of an incorrectly calculated Xpress value. 193

Figure 5. 4 : Analysis of all proteins found in at least 2 experiments using software commonly used in microarray analysis, Babelomics. 205

Figure 5. 5 : Analysis potentially interesting proteins with an average ASAPRatio ≥2.00 using software commonly used in microarray analysis, Babelomics..... 213

Figure 5. 6 : Comparison of proteins identified in RG/C2 and PRG/C2 SILAC proteomic profile with other similar studies to identify secreted proteins in CM derived from various cell lines. 215

Chapter 6 : Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression

Figure 6. 1 : Confirmation of the presence of 3 proteins with an ASAPRatio of 1.00 using Western Blot analysis. 224

Figure 6. 2 : Confirmation of the presence of EMMPRIN in MV's prepared to validate the SILAC results. 226

Figure 6. 3 : Confirmation of the presence of CEACAM5 in MV's prepared to validate the SILAC results. 228

Figure 6. 4 : Confirmation of the presence of AGR2 in MV's prepared to validate the SILAC results. 231

Figure 6. 5 : Confirmation of the presence of OLFM4 in MV's prepared to validate the SILAC results. 233

Figure 6. 6 : Confirmation of the presence of SBP1 in MV's prepared to validate the SILAC results. 234

Figure 6. 7 : Confirmation of the presence of HSP90α in MV's prepared to validate the SILAC results. 238

Figure 6. 8 : Confirmation of the presence of HSP90β in MV's prepared to validate the SILAC results. 239

Figure 6. 9 : Summary of results from the Western blot validation experiments and the comparison with the ASAPRatio from the SILAC experiments. 241

List of Tables

Chapter 1 : Introduction

Table 1. 1: Other Potential Protein Biomarkers for CRC. 29

Table 1. 2 : Potential Biomarkers Identified in MS-based Proteomic Studies. 74

Chapter 2 : Materials and Methods

Table 2. 1 : Sucrose Density Gradient Ultracentrifugation Fractions 104

Table 2. 2 : Antibodies. 112

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify possible biomarkers for the early detection of colorectal cancer using proteomic techniques

Table 3. 1 : The identification of KIAA0051 (gi40789059) in the experiments outlined in Figures 3.7 and 3.8. 152

Chapter 4 : Optimisation of Conditions Required for Quantitative Proteomic Profiling of MVs Released by RG/C2 and PRG/C2 Cell Line

Table 4. 1 : Determination of the sensitivity of the LCQ Deca XP_{PLUS} using BSA digested with trypsin in solution. 158

Table 4. 2 : Various filter parameters were changed to optimise those suitable for the analysis of large datasets using Bioworks Browser v3.1. 168

Table 4. 3 : Which Database? IPI vs NCBI. 174

Table 4. 4 : False positive determination of a single sample following database searching using a decoy database. 179

Table 4. 5 : False positive rate determination of all samples in a single experiment following database searching using a decoy database. 183

Chapter 5 : Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labelling of Amino acids in Culture (SILAC)

Table 5. 1 : Comparison of Xpress ratio and manually calculated ratio. 195

Table 5. 2 : Comparison of the results from the TPP analysis of the 4 SILAC experiments using protein probability values of >0.90 and >0.70.....199

Table 5. 3 : Potentially interesting proteins suitable for validating the results of the SILAC proteomic profile of MV's secreted by RG/C2 and PRG/C2 cell lines.....210

Chapter 6 : Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression

Table 6. 1 : Proteins chosen to validate ASAPRatios calculated by TPP. 221

List of Supplementary Material (see enclosed CD)

Chapter 3 : Investigating the secretome from RG/C2 and PRG/C2 cells to identify possible biomarkers for the early detection of colorectal cancer using proteomic techniques

Supplementary Table 3a: Proteins identified in MV's from CM derived from RG/C2 cells ($6x10^6$ cells) as outlined Figure 3.6.

Supplementary Table 3b: Proteins identified in MV's from CM derived from RG/C2 cells ($1.2x10^7$ cells) as outlined in Figure 3.7.

Supplementary Table 3: Proteins identified in MV's from CM derived from RG/C2 cells ($2.4x10^7$ cells) as outlined in Figure 3.8.

Chapter 5 : Quantitative analysis of proteins secreted in MV's from CM derived from colorectal cancer cell lines using Stable Isotope Labeling of Amino acids in Culture SILAC)

Supplementary Table 5a: Initial analysis of the data generated from the first SILAC experiment using the multiple thresholds filter parameters optimized in table 4.2.

Supplementary Table 5b: Comparison of the TPP results for all 4 SILAC experiments.

Supplementary Table 5c: Proteins identified in these SILAC surveys that were previously reported in proteomic profiles of CM and MVs derived from various cancer cell lines.

Supplementary Table 5d: Raw data from the TPP analysis for experiment C13 ARG A.

Supplementary Table 5e: Raw data from the TPP analysis for experiment C13 ARG B.

Supplementary Table 5f: Raw data from the TPP analysis for experiment C13 LYS A.

Supplementary Table 5g: Raw data from the TPP analysis for experiment C13 LYS B.

Supplementary Table 5h: Proteins identified in these SILAC surveys that were previously reported in a proteomic profile of MVs derived from the HT29 cell line.

Supplementary Table 5i: Proteins identified that were unique to this study when compared to other secreted protein proteomic surveys.

Chapter 6 : Validation of selected proteins identified in the SILAC profile of MV released by colorectal cancer cell lines during early tumour progression

Supplementary Table 6a: Table to show the TPP results for the control proteins.

Supplementary Table 6b: Table to show the TPP results for CEACAM5.

Supplementary Table 6c: Table to show the TPP results for AGR2.

Supplementary Table 6d: Table to show the TPP results for OLFM4.

Supplementary Table 6e: Table to show the TPP results for SBP1.

Supplementary Table 6f: Table to show the TPP results for HSP90α.

Supplementary Table 6g: Table to show the TPP results for HSP90β.

Supplementary Table 6h: Table to show peptide sequences identified that are common to both HSP90α and HSP90β.

Supplementary Table 6i: Table to show peptide sequences identified that are unique to HSP90α.

Supplementary Table 6j: Table to show peptide sequences identified that are unique to HSP90β.

Supplementary Figure 6a: Amino acid sequence of HSC70 and its identified peptides.

Supplementary Figure 6b: Amino acid sequence of TACSTD1 and its identified peptides.

Supplementary Figure 6c: Amino acid sequence of Annexin A1 and its identified peptides.

Supplementary Figure 6d: Amino acid sequence of CEACAM5 and its identified peptides.

Supplementary Figure 6e: Amino acid sequence of AGR2 and its identified peptides.

Supplementary Figure 6f: Amino acid sequence of OLFM4 and its identified peptides.

Supplementary Figure 6g: Amino acid sequence of SBP1 and its identified peptides.

Supplementary Figure 6h: Amino acid sequence of HSP90α and its identified peptides.

Supplementary Figure 6i: Amino acid sequence of HSP90β and its identified peptides.

Abbreviations

2D DIGE - 2D Differential In-Gel Electrophoresis

2D PAGE – Two-Dimensional Polyacrylamide Gel Electrophoresis

AGR2 – Anterior Gradient Homolog 2

AQUA – Absolute Quantification

ASAPR – Automated Statistical Analysis of Protein Abundance Ratios

BSA – Bovine Serum Albumin

CBB G250 - Coomassie Brilliant Blue G250

CEACAM5 (CEA) – Carcinoembryonic Antigen –Related Cell Adhesion Molecule 5

CID - Collision Induced Dissociation

CM – Conditioned media

CRC – Colorectal Cancer

DMEM – Dulbecco's Modified Eagle's Medium

DMSO – Dimethyl Sulphoxide

DTT – Dithiothreitol

EMMPRIN – Extracelleular Matrix Metalloproteinase Inducer

ESI – Electrospray Ionisation

FAP – Familial Adenomatous Polyposis

FCS – Foetal Calf Serum

FOBT - Faecal Occult Blood Test

GO – Gene Ontology

HNPCC – Hereditary Non-Polyposic Colon Cancer

HPLC – High Performance Liquid Chromatography

HSP90 – Heat Shock Protein 90

ICPL – Isotope-Coded Protein Labels

IEF – Isoelectric Focusing

IPG – Immobilised pH Gradients

iTRAQ – Isobaric Tags for Relative and Absolute Quantification

LC – Liquid Chromatography

m/z – mass to charge ratio

MALDI – Matrix Assisted Laser Desorption Ionisation

MRM – Multiple Reaction Monitoring

MV – Microvesicles

OLFM4 – Olfactomedin 4

PBS – Phosphate Buffered Saline

REM CM – Remaining Conditioned media

RP – Reverse Phase

SBP1 – Selenium Binding Protein 1

SCX – Strong Cation Exchange

SELDI – Surface Enhanced Laser Desorption Ionisation

SF – Serum Free

SILAC – Stable Isotope Labelling of Amino Acids in Culture

TACSTD1 – Tumour Associated Calcium Signal Transducer 1

TOF – Time-of-Flight